

Research on Creating an Income Consistent Combined 2014 CPS ASEC File

Jonathan L. Rothbaum, PhD

Expert Meeting on Income, Poverty, and Health Insurance

March 20, 2015

Outline

1. Income sources affected by redesign
2. Using 2014 CPS ASEC for income comparisons
 - Challenges
 - Options
3. Missing information and imputation
 - Creating a revised file that reflects how traditional respondents *would have* answered if given the redesigned questions
4. Results and discussion

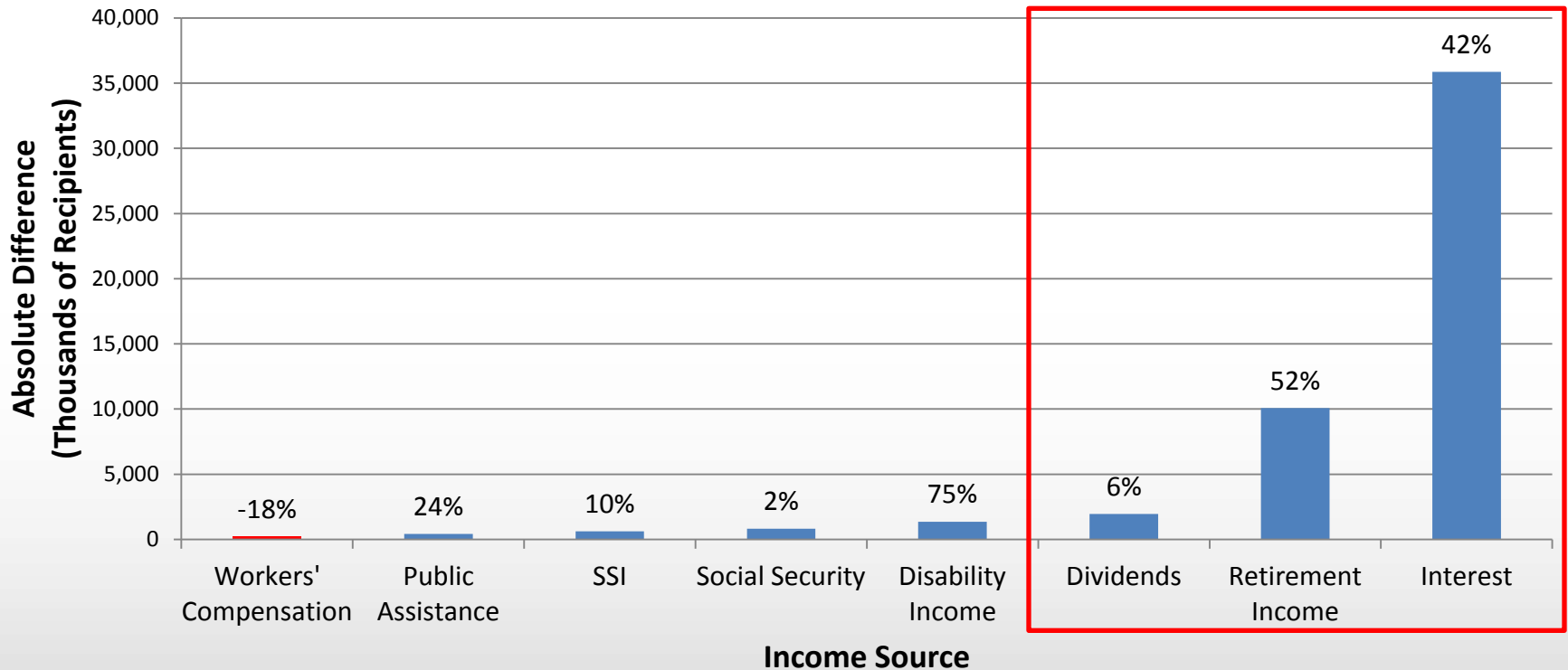
Income Affected by Survey Redesign: No Change in Earnings

Type	Recipients (Thousands)		Average (For Recipients)		Aggregate (Millions)	
	Traditional	Redesign	Traditional	Redesign	Traditional	Redesign
Total Income	218,662	222,135	41,319	42,366	9,035,004	9,410,910
Earnings	158,081	158,571	44,416	44,983	7,021,280	7,133,057

Source: 2014 CPS ASEC, Income for calendar year 2013.

Income Affected by Survey Redesign: Number of Recipients

Absolute Difference in Number of Recipients | Redesign-Traditional |
(% Difference above Bar)

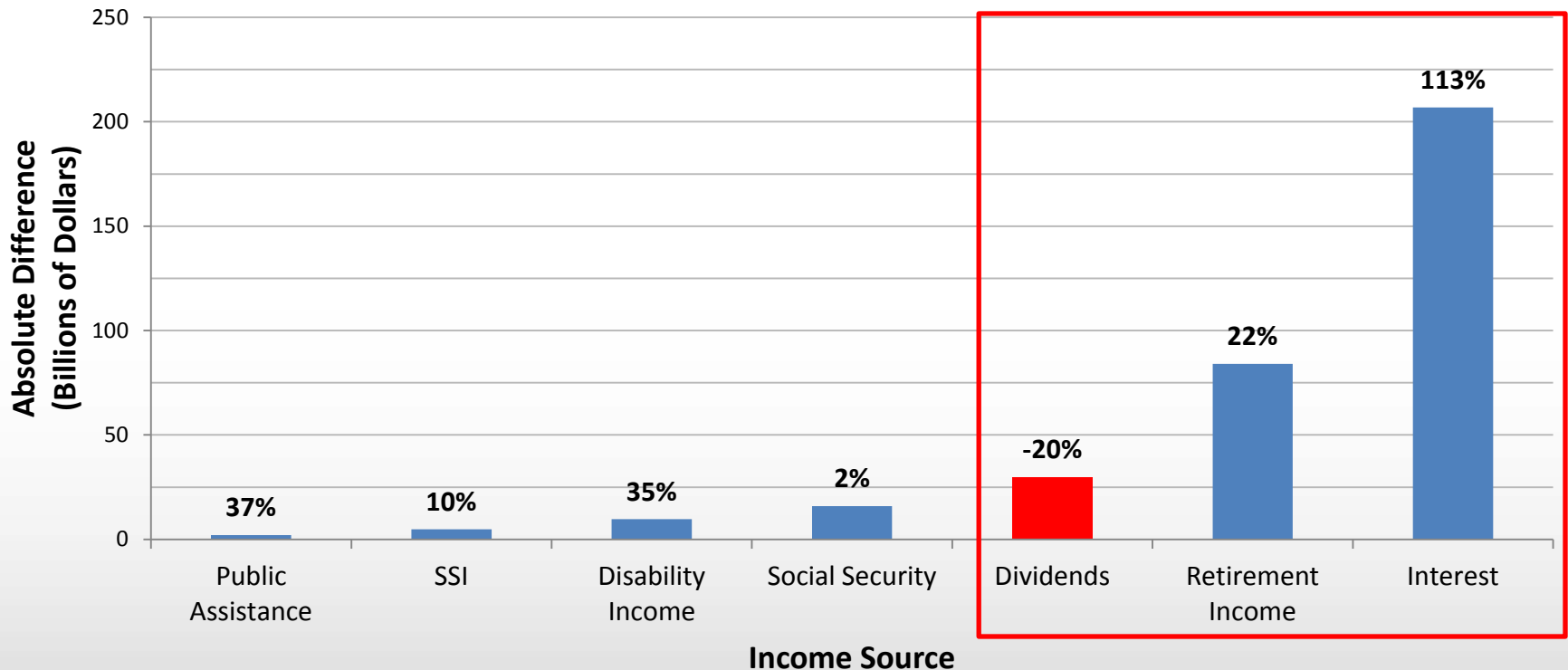


All differences are statistically significant at the 90% level.

Source: 2014 CPS ASEC, Income for calendar year 2013.

Income Affected by Survey Redesign: Aggregate Income

Absolute Difference in Aggregate Income | Redesign-Traditional |
(% Difference above Bar)



All differences are statistically significant at the 90% level.

Source: 2014 CPS ASEC, Income for calendar year 2013.

Challenge

- Survey redesign affected response to a set of income categories, particularly:
 - Retirement Income
 - Interest Income
 - Dividend Income
- Majority of income comes from sources with very little difference between the two samples (89% in Redesign file)

Options

1. Use Redesign file only
 - Wide standard errors
 - Not using any information from Traditional file
2. Combine files without changing Traditional file
 - Ignoring clear differences in responses
3. Treat as a problem of missing information and use imputation
 - Use information from Traditional file that is unaffected by redesign (most income sources, all other variables)
 - No response for changed questions for traditional sample, so impute to estimate what the responses would have been if given the redesigned instrument

Imputation Method

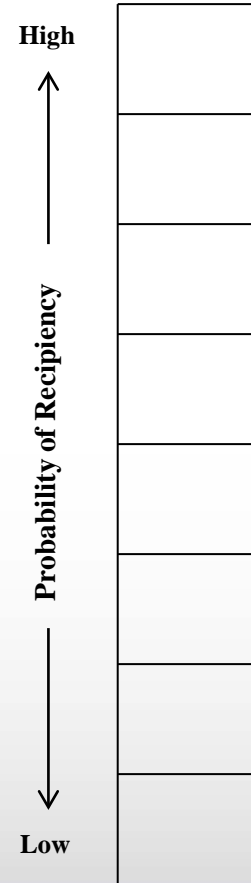
- Use individuals from the Redesign file as donors for individuals in the Traditional file in the affected income categories, as in standard hot deck
- Match individuals along two dimensions:
 1. Probability of reciprocity
 2. Expected income conditional on reciprocity
- Condition on a very rich set of demographic characteristics, income (for non-affected income types), etc.

Step-By-Step Example

1. Predict probability of interest income reciprocity in Redesign for all individuals in both samples

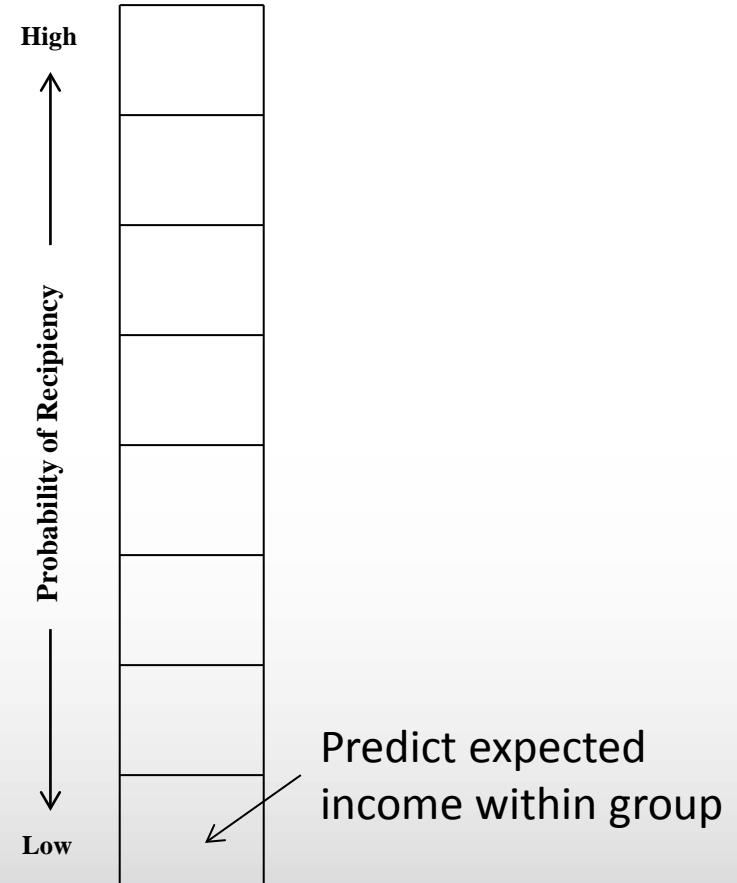
Step-By-Step Example

1. Predict probability of interest income reciprocity in Redesign for all individuals in both samples
2. Divide all individuals in both samples into groups based on this probability



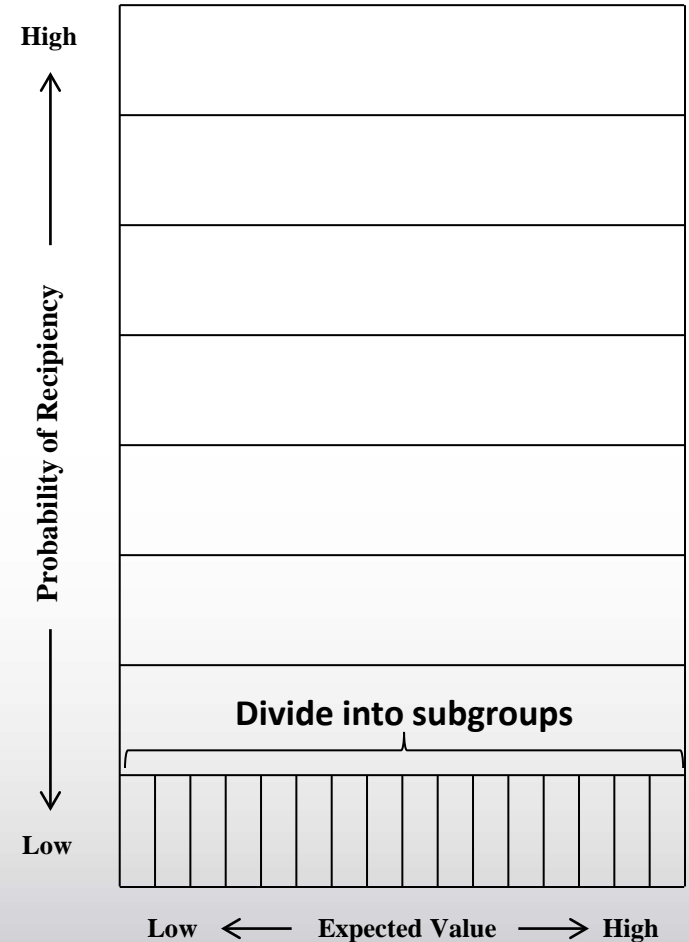
Step-By-Step Example

3. For each group, predict expected interest income in Redesign for all individuals in both samples



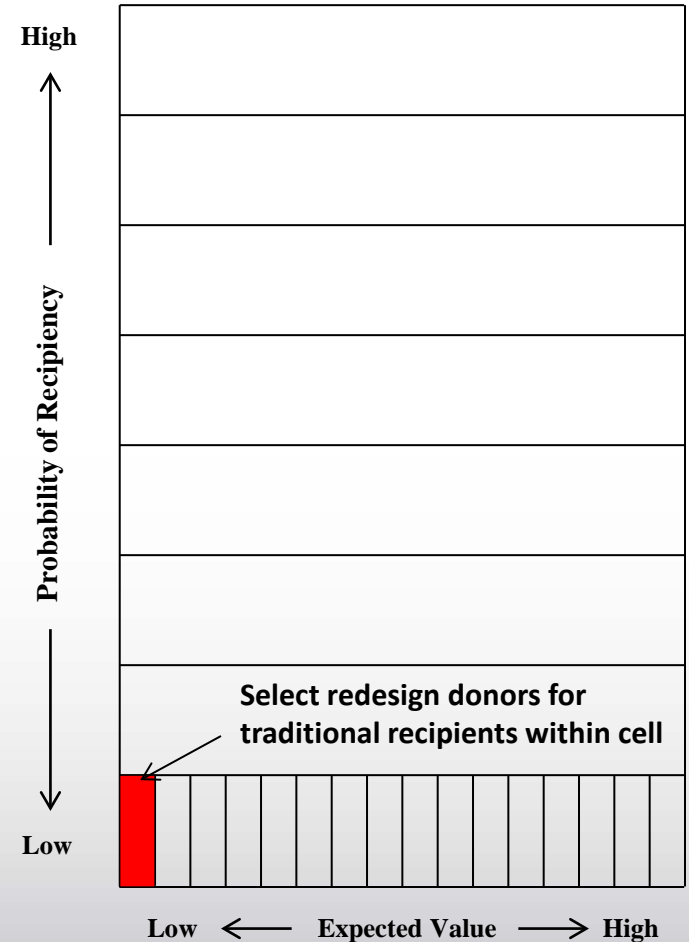
Step-By-Step Example

3. For each group, predict expected interest income in Redesign for all individuals in both samples
4. Divide each group into subgroups based on predicted interest income



Step-By-Step Example

5. Within each subgroup cell, randomly select Redesign individuals as donors for Traditional individuals and donate all variables related to interest income



Advantages of This Approach

- Match donors as in the usual hot deck used in CPS ASEC imputation
- Allows many more variables to be included in model for predicting recency and income than are possible in a hot deck
 - Especially important given number of individuals in smaller Redesign file and number of missing observations (all of Traditional file for the affected income sources)
- Flexibility in modelling for each variable and for each subgroup
- Do not need to specify exact form of model in advance, let data determine appropriate model
- Narrower confidence intervals – smaller changes necessary for statistical significance

Multiple Imputation

- Repeat the process 10 times
- Improve point estimates for statistics such as poverty
 - Average estimates from each completed file
- Account for uncertainty due to imputed values

Processing and Multiple Imputation

- Not possible to release all tables and report data for multiple files
- Which completed file to use in tables and reports?
 - The file with estimates that are closest to the average from the multiple imputation estimates

Choosing the File

- Compute standardized distance of file estimates from average from multiple files
 - Calculate “z-score” based on standard deviation and average of each parameter
 - Find file with minimum sum of squared z-scores
- Which statistics to use?
 - Poverty rate
 - CPS ASEC is source of official poverty rate
 - Median household income
 - Headline statistic
 - Accurate estimate of income necessary for estimate of poverty

Summary of Results

File	Household Median Income	Poverty
Redesign	53,514	14.72%
Traditional	51,939	14.48%
Multiple Imputation		
Income Consistent Average	53,311	14.49%
Closest File to Average (Selected File)	53,314	14.51%

Source: 2014 CPS ASEC, Income for calendar year 2013.

Poverty by Age Group

File	Children (0-17)	Working Age Adults (18-64)	Aged 65 and Older (≥65)
Redesign	21.29%	13.27%	10.30%
Traditional	19.91%	13.56%	9.51%
Multiple Imputation			
Income Consistent Average	20.43%	13.36%	9.65%
Selected File	20.53%	13.36%	9.62%

Source: 2014 CPS ASEC, Income for calendar year 2013.

Comparison of Statistically Significant Changes Using Income Consistent and Redesign Files

File	Estimate of Bounds for Statistically Significant Change (90% CI)			
	Poverty		HH Median Income	
	Lower	Upper	Lower	Upper
Redesign	14.21	15.24	52,371	54,656
Income Consistent Selected File	14.21	14.81	52,628	53,999

Source: 2014 CPS ASEC, Income for calendar year 2013. Estimates based on standard errors from previous CPS ASEC estimates of poverty and median income.

Options

1. Use Redesign file only
 - Wide standard errors
 - Not using any information from Traditional file
2. Treat as a problem of missing information and use imputation
 - Use information from Traditional file that is unaffected by redesign (most income sources, all other variables)
 - No response for changed questions for traditional sample, so impute to estimate what the responses would have been if given the redesigned instrument

Contact Information

Jonathan Rothbaum

Economist, Income Statistics Branch

Social, Economic, and Housing Statistics Division

jonathan.l.rothbaum@census.gov

(301) 763-9681